

'*Candidatus* Phytoplasma vignae', assigning a species description to a long-known phytoplasma occurring in northern Australia

Bianca Rodrigues Jardim^{1,2,*}, Lucy T. T. Tran-Nguyen³, Cherie Gambley⁴, Craig Webster⁵, Monica Kehoe⁵, Samantha Bond⁶, Brendan Rodoni^{1,2} and Fiona E. Constable^{1,2}

Abstract

Gene- and genome-based approaches were used to determine whether *Vigna* little leaf (ViLL) phytoplasma, which occurs in northern Australia, is a distinct '*Candidatus* Phytoplasma' species. The ViLL 16S rRNA gene sequences exhibited the highest known similarity to species in the 16SrXXIX-A and 16SrIX-D subgroups, namely '*Candidatus* Phytoplasma omanense' (98.03–98.10%) and '*Candidatus* Phytoplasma phoenicium' (96.87–97.20%), respectively. A total of 48 single-copy orthologue genes were identified to be shared among the two draft ViLL phytoplasma genomes, 30 publicly available phytoplasma genomes, and one *Acholeplasma laidlawii* genome as the outgroup taxon. Phylogenomic assessments using the 48 shared single-copy orthologue genes supported that ViLL and '*Ca.* Phytoplasma phoenicium' were closely related yet distinct species. The 16S rRNA gene sequence analysis and phylogenomic assessment indicate that ViLL phytoplasmas are a distinct taxon. As such, a novel species, '*Candidatus* Phytoplasma vignae', is proposed. Strain BAWM-336 (genome accession number JAUZLI000000000) detected in *Momordica charantia* (bitter melon) serves as the reference strain of this species, with infected plant material deposited in the Victorian Plant Pathology Herbarium (VPRI) as VPRI 44369.

INTRODUCTION

The *Vigna* little leaf (ViLL) phytoplasma has historically been reported in the Northern Territory, Australia, near the townships of Katherine and Darwin (Fig. 1). It was named for the little leaf symptoms associated with its infection of *Vigna lanceolata* (pencil yam, family: *Fabaceae*), first detected in Katherine [1]. Based on Sanger sequencing and/or restriction fragment length polymorphism (RFLP) analyses, the ViLL phytoplasma was subsequently found associated with phyllody in *Tridax procumbens* (tridax daisy, family: *Asteraceae*) also collected near Katherine [2], and little leaf symptoms in *Stylosanthes scabra* (shrubby stylo, family: *Fabaceae*) from Darwin [3]. In a few instances, mixed infections of both ViLL and *Stylosanthes* little leaf (StLL) phytoplasmas (now described as '*Candidatus* Phytoplasma stylosanthis' [4]) were reported [3, 5].

In 2021, ViLL phytoplasmas were detected in a single *Catharanthus roseus* (periwinkle, family *Apocynaceae*) plant showing symptoms of yellowing and little leaves sampled from a private garden in Kununurra, Western Australia [6], and in *Momordica charantia* (bitter melon, family *Cucurbitaceae*) sampled from a single symptomatic plant collected from a field in Marrakai, Northern Territory (Fig. 1). Disease in this bitter melon crop was unusually high, with between 70 and 80% of

Author affiliations: ¹School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia; ²Agriculture Victoria Research, Department of Energy, Environment and Climate Action, AgriBio, Bundoora, Australia; ³Plant Health Australia, Deakin, Australian Capital Territory, Australia; ⁴Horticulture and Forestry Science, Department of Agriculture and Fisheries Maroochy Research Facility, Nambour, Australia; ⁵Diagnostic Laboratory Services, Department of Primary Industries and Regional Development, South Perth, WA, Australia; ⁶Biosecurity and Animal Welfare, Department of Industry, Tourism and Trade, Darwin, Australia.

***Correspondence:** Bianca Rodrigues Jardim, 20171100@students.latrobe.edu.au; bianca.rodriguesjardim@agriculture.vic.gov.au; b.rodriguesjardim.agric@gmail.com

Keywords: *Mollicutes*; phylogenomics; phytopathogen; ViLL; *Vigna* little leaf phytoplasma.

Abbreviations: gCF, gene concordance factor; IRPCM, International Research Programme on Comparative Mycoplasma; RFLP, restriction fragment length polymorphism; sCF, site concordance factor; StLL, *Stylosanthes* little leaf; ViLL, *Vigna* little leaf.

The genomes of '*Candidatus* Phytoplasma vignae' were submitted to the NCBI under the genome accession numbers JAUZJL000000000 (strain BAWM-245) and JAUZLI000000000 (strain BAWM-336) as part of the BioProj PRJNA980440. The 16S rRNA gene sequences of each strain were deposited at the NCBI with the following accession numbers PP575759 (BAWM-245) and PP575760 (BAWM-336). Preserved plant material of samples BAWM-245 and BAWM-336 were submitted to the Victorian Plant Pathology Herbarium (VPRI) under the accessions VPRI 44368 and VPRI 44369, respectively.

Two supplementary tables are available with the online version of this article.

006502 © 2024 The Authors



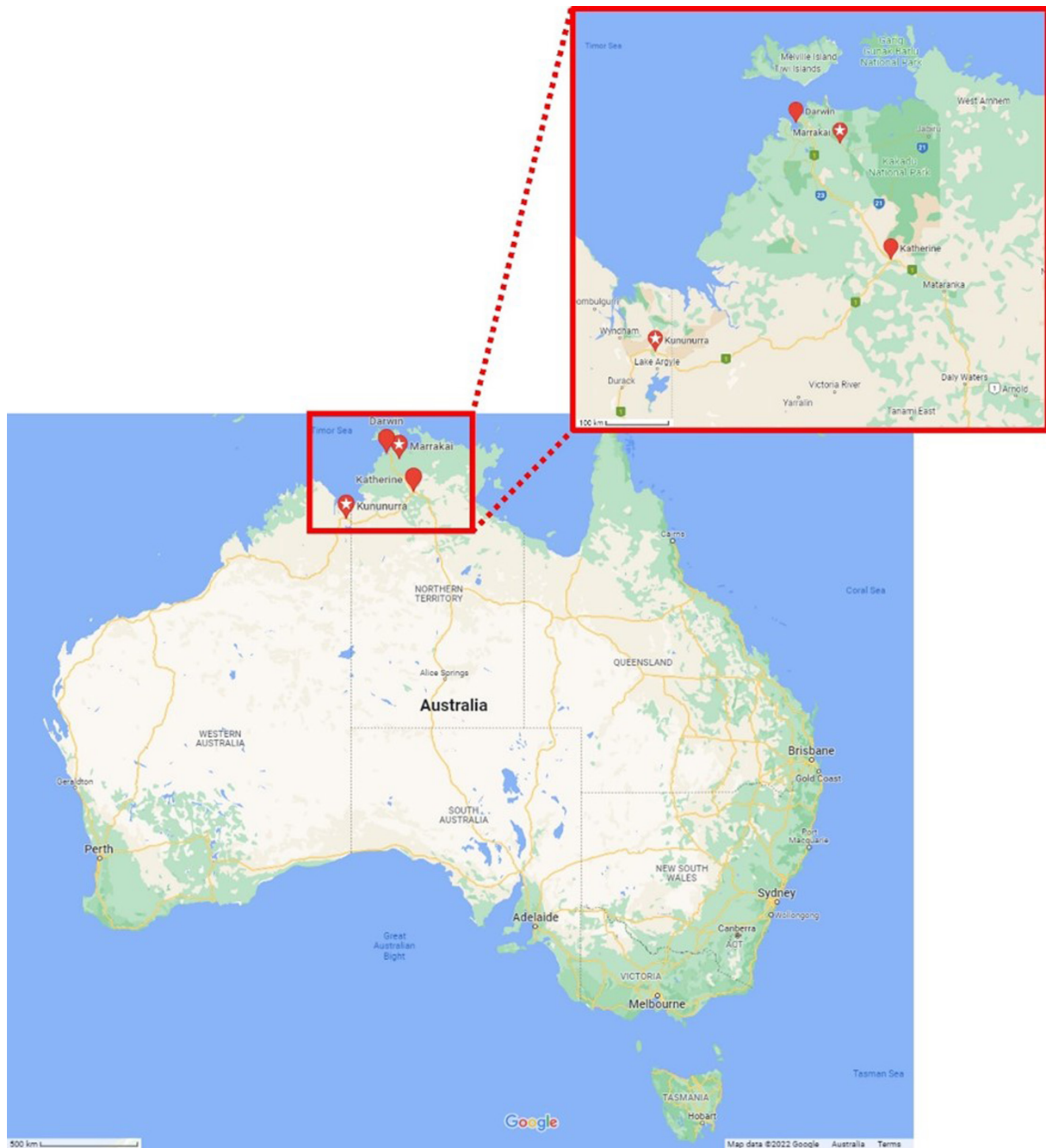


Fig. 1. Approximate geographic locations of *Vigna* little leaf detections in Australian surveys. Solid red pins indicate historical detections according to [1–3], while pins with a white centre star indicate the locations of BAWM-245 and BAWM-336 used in this study.

the *M. charantia* crop showing symptoms of phyllody, distorted fruit shape, and witches' broom (Samantha Bond, personal communication). These ViLL phytoplasma detections made in 2021 represent a new report for Western Australia and expand the known host range of the phytoplasma into two new plant host families, namely the families *Apocynaceae* and the *Cucurbitaceae*. Based on the recent and historical detections, this phytoplasma may be restricted to the northern regions of Australia and may have a diverse host range. However, the ViLL phytoplasma remains an infrequently detected taxon in field surveys conducted in Australia [1, 5, 6].

Phytoplasmas are unculturable microbes and, as such, the phenotypic properties of species cannot be assessed directly. The polyphasic approach of combining phenotypic and genealogical traits in a species description are therefore not possible for this group of bacteria. Instead, genealogical evidence and biological properties, when available, are used to delimit a novel phytoplasma

Table 1. Sample details of two ViLL phytoplasmas analysed in this study including sample name, host species, symptoms and sampling location in Australia as well as DNA paired-end metagenomic sequence data quality evaluations and genome assembly statistics

Data quality evaluations include the percent of raw reads passing filter and total reads per library per sample. Genome assembly statistics include sequence depth coverage, G+C content and genome size of ViLL phytoplasma genomes assembled and analysed in this study for phylogenetic and phylogenomic evaluations.

Details	BAWM-245	BAWM-336
Victorian Plant Pathology Herbarium (VPRI) accession no.	VPRI 44368	VPRI 44369
Genome accession number	JAUZJL000000000	JAUZLI000000000
Host common name (species name)	Periwinkle (<i>Catharanthus roseus</i>)	Bitter melon (<i>Momordica charantia</i>)
Host symptoms	Leaf yellowing, mild leaf size reduction, mild proliferation	Phyllody, distorted fruit shape, witches'-broom
Sampling location	Kununurra, Western Australia	Marrakai, Northern Territory
Total reads before QC (million)	7.0	14.61
Total reads after QC (million)	6.9	14.46
Genome size (bp)	540232	586186
Contigs >1000 bp (<i>n</i>)	34	28
Coverage (×)	233.0	99.3
N50 (bp)	29777	76363
G+C content (mol%)	23.4	24.0
tRNA genes (<i>n</i>)	33	32
rRNA genes (<i>n</i>)	1	1
Protein-coding genes (<i>n</i>)	459	472

species [7, 8]. According to the Phytoplasma Taxonomy Group within the International Research Programme for Comparative Mycoplasma (IRPCM), a novel '*Ca. Phytoplasma*' taxon can be described when the 16S rRNA gene sequence bears less than 98.65% nucleotide sequence similarity to previously described '*Ca. Phytoplasma*' taxa [8]. In cases where the 16S rRNA gene sequence is more than 98.65% similar to a previously described taxon but has a distinct vector and host range or is associated with distinct symptomology in the same host, and there is evidence of significant molecular or serological diversity, a novel taxon can still be assigned. The use of either multilocus sequence analysis or genome sequence comparisons are recommended and implemented in novel species descriptions [9–11], especially when the 16S rRNA gene sequence similarity is higher than the 98.65% threshold [8].

SEQUENCE ANALYSES OF THE VILL PHYTOPLASMAS

Although previous studies indicated that the ViLL phytoplasma is a distinct taxon based on the 16S rRNA gene sequence [1–3, 5], it has not been characterized as a novel '*Ca. Phytoplasma*' species and is therefore not considered in phylogenetic analyses. Here, we investigate the phylogenetic relationship of ViLL phytoplasmas detected in symptomatic periwinkle and bitter melon plants using 16S rRNA nucleotide sequence similarity and phylogenetic analyses, as well as a phylogenomic approach (Table 1).

Total DNA was extracted from symptomatic petiole and leaf tissue, and the DNA quality and quantity assessed according to [4]. All total DNA extractions were stored at –20°C until use. Total DNA extracts were used in a nested PCR using P1/P7 primers followed by R16F2n/m23sr primers for 16S rRNA-based phytoplasma identification [12] and sent for Sanger sequencing after size confirmation by agarose gel electrophoresis according to [4]. Each amplicon sequence per sample was assembled into a consensus sequence using Geneious Prime 2022.1.1 (www.geneious.com) and submitted for BLASTn analyses [13] at the NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Both assembled amplicons showed over 99.7% sequence similarity to the 16S rRNA gene from previously submitted ViLL phytoplasmas (Genbank accession numbers: Y15866 and AJ289195), indicating detections of ViLL phytoplasmas in the periwinkle and bitter melon hosts. The phytoplasma-positive plant material was preserved by freeze

drying in individual screw cap tubes for at least 72 h at -50°C using the FreeZone 2.5l Benchtop Freeze Dry System (Labconco) and submitted to the Victorian Plant Pathology Herbarium (VPRI).

Illumina high-throughput sequencing libraries were prepared, quantified, and sequenced on the Illumina NovaSeq 6000 platform with the SP Reagent Kit version 1.5 as described previously [14]. FastP [15] was used to discard sequences shorter than 50 bp and to remove adapters and sequences below the Q20 threshold. Approximately 6.9 million reads remained for BAWM-245 (98.59% reads passing filters; Table 1), and 14.46 million reads remained for BAWM-336 (98.91% reads passing filters; Table 1). Trimmed reads were assembled and phytoplasma-derived contigs identified and retrieved according to [14]. Retrieved phytoplasma-derived contigs that were shorter than 1000 bp were removed from the assembly. The remaining phytoplasma-derived contigs were submitted to metaQUAST [16] to estimate the G+C content, genome size and N50 values. The phytoplasma genome assembly from sample BAWM-336 was the largest of the two assemblies at 586186 bp, with the largest N50 value (76363 bp) (Table 1). The phytoplasma genome assembly from sample BAWM-245 was smaller at 540232 bp with an N50 value of 29777 bp (Table 1). The G+C content was 23.4 and 24.0 mol% for BAWM-245 and BAWM-336, respectively. Each of the phytoplasma genome sequences were also used as the reference in BBSplit 38.61 within the BBMap version 38.61b software suite [17] with the corresponding trimmed reads used as input to estimate the coverage of each genome. The genome sequence of BAWM-245 and BAWM-336 had high coverage at $233.0\times$ and $99.3\times$, respectively. Prokka 1.14.5 [18] was used to predict and count the number of protein coding genes, tRNA genes, and rRNA genes per genome, specifying RNAmmer for rRNA prediction [19]. For the BAWM-245 draft phytoplasma genome, 33 tRNA genes, one rRNA gene, and 459 protein-coding genes were annotated. For the BAWM-336 draft genome, 32 tRNAs, one rRNA gene, and 472 protein-coding genes were annotated (Table 1). Together, the genome assembly statistics and gene annotations above suggest sufficient data and contiguity of the draft ViLL phytoplasma genome sequences for use in comparative genomic and phylogenomic assessments [11, 20].

EVIDENCE OF SUFFICIENT MOLECULAR DIVERGENCE TO ASSIGN A NOVEL 'CANDIDATUS PHYTOPLASMA' TAXON

16S rRNA gene sequence analysis

The 16S rRNA genes were extracted from the genome assemblies according to length recommendations specified by [8]. For the extraction, the 16S rRNA sequences corresponding to representative 'Ca. Phytoplasma' taxa were compiled into a reference set of sequences [8]. The reference set was mapped back to the ViLL phytoplasma genome assemblies to extract the gene regions using BBMap [17] implemented in Geneious prime 2022.1.1 (www.geneious.com). The extracted gene regions were submitted for BLASTn analysis [13] to confirm their identities for each sample. The extracted 16S rRNA genes were aligned using MAFFT [21] with previously described 'Ca. Phytoplasma' representatives to estimate the sequence similarities and phylogenetic position of the ViLL phytoplasmas. Pairwise sequence similarities between the two ViLL phytoplasma 16S rRNA gene sequences extracted from the genome assemblies as well as the original sequences deposited on the NCBI database shared 99.80–99.87% nucleotide sequence identity with each other (Table 2). The ViLL phytoplasma 16S rRNA gene sequences exhibited the next highest sequence similarity to 'Ca. Phytoplasma omanense' strain IM-4 (98.03%–98.10%), 'Ca. Phytoplasma phoenicium' strain SA213 (97.12%–97.20%), and 'Ca. Phytoplasma phoenicium' strain A4 (96.87%). However, the 16S rRNA gene sequences of 'Ca. Phytoplasma omanense' strain IM-4 and 'Ca. Phytoplasma phoenicium' strain SA213 are 90 and 264 bp shorter than the recommended length, respectively, suggesting that the percent identities might differ should full-length 16S rRNA gene sequences for

Table 2. Pairwise sequence similarity analyses among the *Vigna* little leaf (ViLL) phytoplasma strains BAWM-245 and BAWM-336 as well as strains of their close relatives, namely 'Candidatus Phytoplasma phoenicium' strain SA213 and 'Candidatus Phytoplasma omanense' strain IM-4

Sequence similarity analyses of the 16S rRNA gene was done for both ViLL strains and their two close 'Candidatus Phytoplasma' relatives.

Samples in pairwise comparison	16S rRNA gene sequence similarity (%)
BAWM-336 vs BAWM-245	99.80
BAWM-336 vs 'Ca. Phytoplasma phoenicium' strain SA213	97.20
BAWM-245 vs 'Ca. Phytoplasma phoenicium' strain SA213	97.12
BAWM-336 vs 'Ca. Phytoplasma phoenicium' strain A4	96.87
BAWM-245 vs 'Ca. Phytoplasma phoenicium' strain A4	96.87
BAWM-336 vs 'Ca. Phytoplasma omanense' strain IM-4	98.03
BAWM-245 vs 'Ca. Phytoplasma omanense' strain IM-4	98.10

these strains become available. According to the phytoplasma species description guidelines published in 2022 [8], these results support the description of ViLL phytoplasma as a novel species as they exhibit less than 98.65% sequence similarity to the 16S rRNA gene of any previously described ‘*Ca. Phytoplasma*’ species.

Alignments of the 16S rRNA genes were used to reconstruct a maximum-likelihood tree using RAxML version 4 implemented using Geneious Prime 2022.1.1 to infer the phylogenetic position of the ViLL phytoplasma strains (bootstrap method with 1000 bootstrap replicates [22]). The phylogenetic analyses of the 16S rRNA gene indicate that the ViLL phytoplasmas form their own, well-supported clade (bootstrap support of 99%), with a close evolutionary relationship to strains of ‘*Ca. Phytoplasma omanense*’ and ‘*Ca. Phytoplasma phoenicium*’ (Fig. 2).

Phylogenomic comparisons

Scaffolds of non-phytoplasma origin were identified in the publicly available draft genome sequence of ‘*Ca. Phytoplasma phoenicium*’ strain SA213 (GenBank accession no.: NZ_JPSQ00000000.1) using BLASTN analyses in previous assessments (data not shown). To ensure accurate comparisons between the genomes analysed in the study, the non-phytoplasma scaffolds were removed from the assembly and used in all further downstream analyses. The removal of the non-phytoplasma scaffolds reduced the ‘*Ca. Phytoplasma phoenicium*’ genome size from 345965 to 312035 bp and reduced the number of protein coding genes from 314 to 275. The ‘*Ca. Phytoplasma phoenicium*’ genome was further determined to be highly incomplete due to the low number of tRNA genes recovered from the genome (three tRNA genes in both the contaminated and contaminated-removed genomes).

Phylogenomic comparisons between the two ViLL genomes, 30 publicly available phytoplasma genomes, and one *Acholeplasma laidlawii* genome (Table S1, available in the online version of this article) were performed as previously described [11]. Briefly, this approach implements Orthofinder 2 [23, 24] to identify the single-copy orthologue genes shared between all genomes considered. A total of 48 shared single-copy orthologue genes were identified among the 33 genomes used in the analyses. The function of the single copy orthologue genes obtained by Prokka annotation are listed in Table S2. The 48 shared single-copy orthologue genes were individually aligned using MAFFT [21], and then analysed in IQ-TREE version 2.0.5 [25–27] to concatenate and construct the subsequent phylogenomic tree. Branch support was shown using gene concordance factor (gCF) and site concordance factor (sCF) values, with gCF and sCF values of $\geq 50\%$ indicating high statistical support.

The high gCF (100%) and sCF (97.23%) values of the ViLL branch in the phylogenomic tree and their short respective branch lengths support that the two strains are closely related. Additionally, the branch encompassing the two ViLL phytoplasma strains and ‘*Ca. Phytoplasma phoenicium*’ strain SA213 show high gCF (97.92%) and sCF (58.94%) support values supporting that the ViLL strains are distinct species from ‘*Ca. Phytoplasma phoenicium*’ (Fig. 3). Unfortunately, no genome sequences were available for comparison with the known ViLL relative, ‘*Ca. Phytoplasma omanense*’ at the time of this study. Additionally, the genome of ‘*Ca. Phytoplasma phoenicium*’ strain SA213 is highly incomplete and therefore limited the number of single-copy orthologues obtained for the phylogenomic analyses (Table S2).

DESCRIPTION OF ‘*CANDIDATUS PHYTOPLASMA VIGNAE*’

‘*Candidatus Phytoplasma vignae*’ (vig’nae N.L. gen. n. *vignae*, of *Vigna*, referring to the plant genus in which the phytoplasma was first detected).

The reference strain is BAWM-336^R associated with *Momordica charantia* (bitter melon) showing symptoms of phyllody, fruit shape distortions, and witches'-broom. [Mollicutes] NC; NA; O, wall-less; NAS (GenBank accession number: JAUZLI000000000); G+C content 24.0 mol%; oligonucleotide sequences of unique regions of the 16S rRNA gene are shown in Table 3.

Preserved *M. charantia* (bitter melon) tissue of sample BAWM-336^R from Marrakai, Northern Territory, Australia, was deposited at the Victorian Plant Pathology Herbarium (VPRI) under the accession VPRI 44369 at Agriculture Victoria, Bundoora, Australia. Sample BAWM-245 was also submitted to the VPRI (accession VPRI 44368). We propose that the genome sequence of strain BAWM-336^R (GenBank assembly number: JAUZLI000000000) be considered as a reference genome for ‘*Ca. Phytoplasma vignae*’. This strain was selected as the reference strain because it originated from the Northern Territory where ViLL phytoplasmas were first reported [1]. The 16S rRNA nucleotide sequence analyses also indicate a close relationship of strain BAWM-336^R with that of previous detections in the Northern Territory (Fig. 2).

The results presented here provide multiple lines of evidence using more than one sample to support the description of ViLL as a novel phytoplasma species, including comparisons of gene sequence identities (16S rRNA gene) and reconstructing phylogenies (16S rRNA gene and shared single-copy orthologues). These analyses are in-line with the updated recommendations for phytoplasma species descriptions [8] and are an improvement over the traditional approaches that only considered the 16S rRNA gene sequence similarity and RFLP [7], which represent a very small fraction of the genome and are much less informative for establishing taxonomic relationships among ‘*Ca. Phytoplasma*’ species.

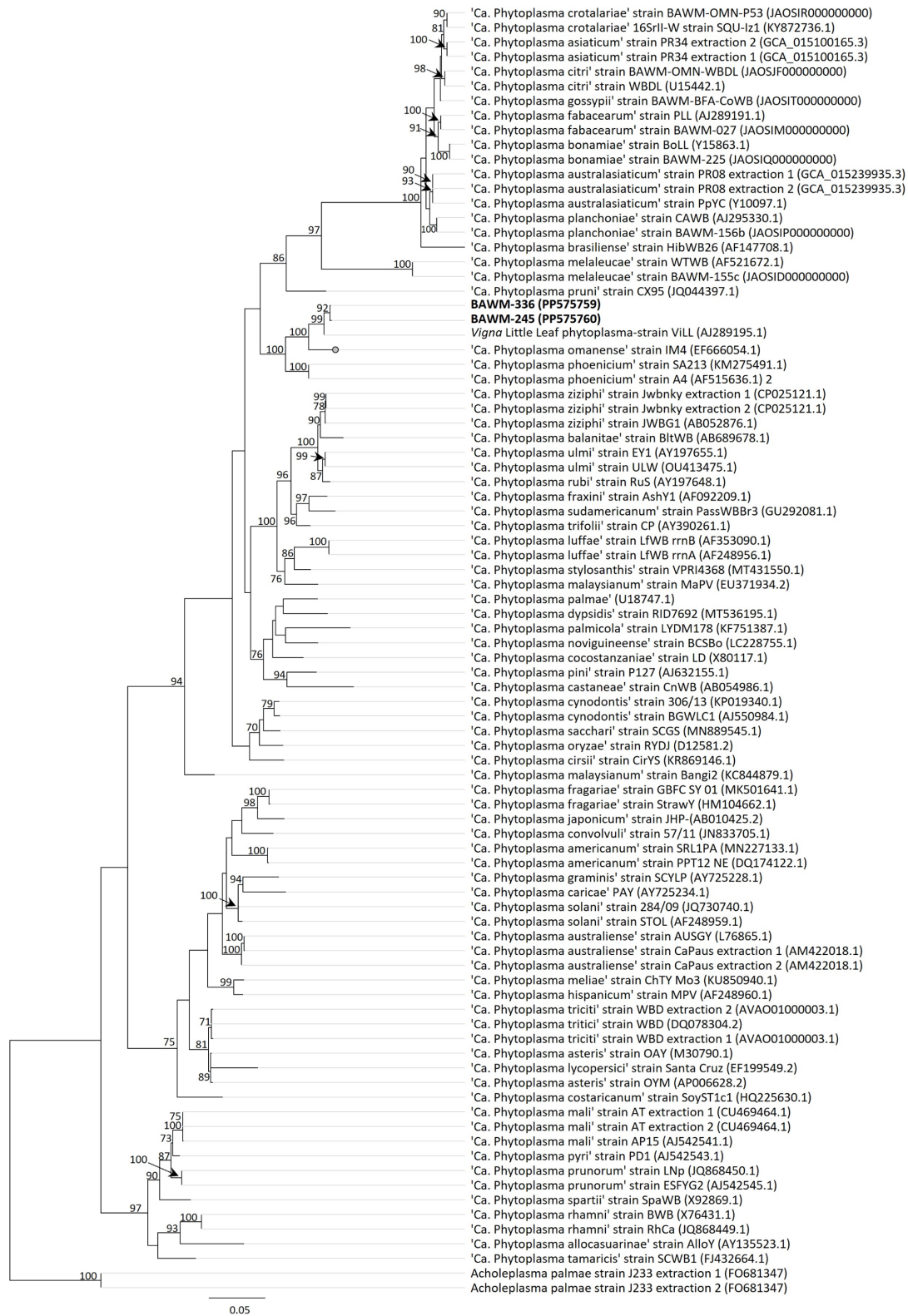


Fig. 2. Maximum-likelihood tree inferred from analysis of the 16S rRNA gene sequences of two *Vigna* little leaf phytoplasma strains (highlighted in bold font) and previously described '*Candidatus* Phytoplasma' reference and related strains. *Acholeplasma palmae* was used as the outgroup taxon. Maximum-likelihood analyses were performed using RaxML implemented in Geneious Prime 2022.1.1. Bootstrap support values >70% are indicated at branch nodes, and branch lengths indicate the number of nucleotide substitutions per site (see bar).

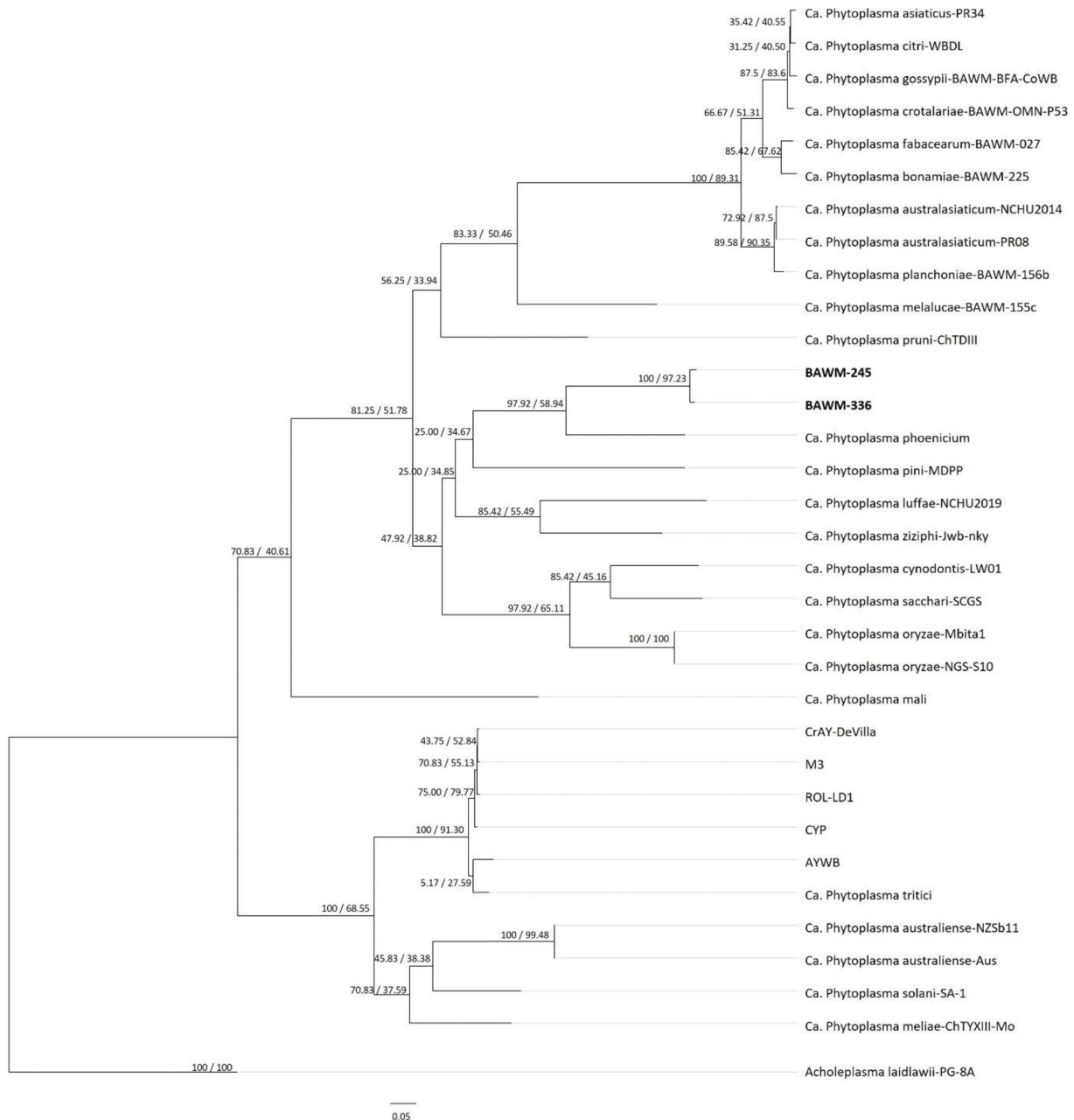


Fig. 3. Maximum-likelihood tree inferred using IQ-TREE version 2.0.5 using the concatenated dataset of 48 single copy orthologue genes identified to be shared across each of the 33 genomes analysed in this study. Support values include gene (gCF) and site concordance factors (sCF), i.e., gCF / sCF. Branch lengths indicate the number of amino acid substitutions per site (see bar). Bold font indicates the *Vigna* little leaf (ViLL) phytoplasma strains analysed in this study.

The genomes obtained in this study serve as an essential resource to aid in improving our understanding of phytoplasma evolution, supporting the epidemiological studies of this species, and advancing phytoplasma diagnostics going forward. The preserved material deposited at the VPRI through this study serves as an important resource that should be revisited to improve the quality and completeness of the genome assemblies of these strains in the future. Putative vectors of ViLL were determined to be *Austroagallia torrida* (Evans) and a *Batracomorphus* species based on direct PCR analyses only, with transmission by these vectors not yet tested [2]. More work is needed to understand the vector, host, and geographic range of '*Ca. Phytolasma vignae*'. Further work to generate high quality genome sequence data from the close known relatives '*Ca.*

Table 3. Unique oligonucleotide sequences and their position in the 16S rRNA gene of '*Ca. Phytoplasma vignae*' reference strain BAWM-336^R

Position	Unique sequence (5' to 3' orientation)
61–137bp	ACCACGAAAGTTGGCAATACCCAAAAGCGGTCCCTAACTTCTTCGGAGAGGGAGCCCGTCTAAGGTAGGTCGATG
301–306bp	ACTGTG
358–453bp	CCTAAAACGAGCGGAACCCCTTATCGTTAGTTGGGACCACGTAATGGTGAGCACCTTTAGCCGAGACTGCCAATGATAAAATTTGGAGGAAGGTGAGGATT
507–583bp	GAAGATACACGAAAAACCTTACCAGGTCTTGACATAATTTGCCGAAAGTTATAGAAAATATAACGGAGGTTATCAGAAT
569–633bp	TTTGGTAAGTCTATAAGTTTAAATTCAGCGCTTAAACGCTGTTGCCGCTATAGAAAACCTGCCTCACTA
1271–1323bp	AAGGTATCTTAAAGAGGGGCTTGGCTCACATTAGTTAGTTGGTAAAGGTAATG

Phytoplasma phoenicium' and 'Ca. Phytoplasma omanense' will improve the understanding of their taxonomic relationship to 'Ca. Phytoplasma vignae'.

Funding information

This project was funded by Hort Innovation using vegetable industry levies with co-investment from the Queensland Department of Agriculture and Fisheries; Victorian Department of Jobs, Precincts and Regions; the Northern Territory Department of Industry, Tourism and Trade; the Western Australia Department of Primary Industries and Regional Development and the University of Tasmania. Hort Innovation is the grower-owned, not-for-profit research and development corporation for Australian horticulture. We would also like to acknowledge and thank Agriculture Victoria Research and La Trobe University for financial support of BRJ through their provision of a Living Allowance Scholarship and the La Trobe Full Fee Research Scholarship (LTUFFRS), respectively.

Acknowledgements

We would like to acknowledge and thank Tracie Webster (Agriculture Victoria Research) for handling the submission and NovaSeq runs of the Illumina libraries generated in this study, as well as Simone Kreidl for her assistance with lodging the preserved plant material to the Victorian Plant Pathology Herbarium (VPRI). This research was conducted using the facilities of Agriculture Victoria Research.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Schneider B, Padovan A, De La Rue S, Eichner R, Davis R, *et al.* Detection and differentiation of phytoplasmas in Australia: an update. *Aust J Agric Res* 1999;50:333.
- Padovan AC, Gibb KS. Epidemiology of phytoplasma diseases in papaya in northern Australia. *J Phytopathol* 2001;149:649–658.
- De La Rue S, Padovan A, Gibb K. *Stylosanthes* is a host for several phytoplasmas, one of which shows unique 16S–23S intergenic spacer region heterogeneity. *J Phytopathol* 2001;149:613–619.
- Rodrigues Jardim B, Kinoti WM, Tran-Nguyen LTT, Gambley C, Rodoni B, *et al.* 'Candidatus Phytoplasma stylosanthis', a novel taxon with a diverse host range in Australia, characterised using multi-locus sequence analysis of 16S rRNA, *secA*, *tuf*, and *rp* genes. *Int J Syst Evol Microbiol* 2021;71:1–11.
- Streten C, Gibb KS. Phytoplasma diseases in sub-tropical and tropical Australia. *Austral Plant Pathol* 2006;35:129–146.
- Rodrigues Jardim B, Gambley C, Tran-Nguyen LTT, Webster C, Kehoe M, *et al.* A metagenomic investigation of phytoplasma diversity in Australian vegetable growing regions. *Microb Genom* 2024;10:001213.
- Firrao G, Andersen M, Bertaccini A, Boudon E, Bové JM. 'Candidatus Phytoplasma', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. *Int J Syst Evol Microbiol* 2004;54:1243–1255.
- Bertaccini A, Arocha-Rosete Y, Contaldo N, Duduk B, Fiore N, *et al.* Revision of the 'Candidatus Phytoplasma' species description guidelines. *Int J Syst Evol Microbiol* 2022;72.
- Kirdat K, Tiwarekar B, Thorat V, Sathe S, Shouche Y, *et al.* 'Candidatus Phytoplasma sacchari', a novel taxon - associated with Sugarcane Grassy Shoot (SCGS) disease. *Int J Syst Evol Microbiol* 2021;71:1–7.
- Zhao Y, Wei W, Davis RE, Lee IM, Bottner-Parker KD. The agent associated with blue dwarf disease in wheat represents a new phytoplasma taxon, "Candidatus Phytoplasma tritici." *Int J Syst Evol Microbiol* 2021;71:1–11.
- Rodrigues Jardim B, Tran-Nguyen LTT, Gambley C, Al-Sadi AM, Al-Subhi AM, *et al.* The observation of taxonomic boundaries for the 16SrII and 16SrXXV phytoplasmas using genome-based delimitation. *Int J Syst Evol Microbiol* 2023;73.
- Constable FE, Joyce PA, Rodoni BC. A survey of key Australian pome fruit growing districts for exotic and endemic pathogens. *Austral Plant Pathol* 2007;36:165–172.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
- Rodrigues Jardim B, Tran-Nguyen LTT, Gambley C, Rodoni B, Constable FE. Iodixanol density gradients as an effective phytoplasma enrichment approach to improve genome sequencing. *Front Microbiol* 2022;13:1–16.
- Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.
- Mikheenko A, Saveliev V, Gurevich A. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* 2016;32:1088–1090.
- Bushnell B. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. 2014.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
- Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007;35:3100–3108.
- Hedlund BP, Chuvochina M, Hugenholtz P, Konstantinidis KT, Murray AE, *et al.* SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat Microbiol* 2022;7:1702–1708.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
- Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
- Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 2015;16.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 2019;20:1–14.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;37:1530–1534.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
- Chernomor O, Von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol* 2016;65:997–1008.